

論文 / 著書情報
Article / Book Information

論題(和文)	区分線形変換による雑音適応法における木構造クラスタリングの検討
Title(English)	
著者(和文)	張 志鵬, 大辻 清太, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2003年春季講演論文集, Vol. , No. 1-4-16, pp. 37-38
Citation(English)	, Vol. , No. 1-4-16, pp. 37-38
発行日 / Pub. date	2003, 3

区分線形変換による雑音適応法における木構造クラスタリング法の検討*

◎張 志鵬 大辻 清太 (NTTドコモ) 古井 貞熙 (東工大)

1. はじめに

大語彙連続音声認識における問題の一つとして、背景に雑音や音楽を含む音声に対する認識性能の劣化が挙げられる。これまでに我々は区分線形変換(PLT; piecewise linear transformation)による雑音適応法における木構造クラスタリング法[1]を提案した。手法[1]では、雑音GMMに対し木構造クラスタリングを行った。この論文では雑音重畳音声GMMによるクラスタリング手法を提案する。また数量化第四類分析により雑音クラスタリングの効果を検討した。

2. 区分線形変換雑音適応法における木構造クラスタリング

2.1 区分線形変換による雑音適応手法

一般に、音声に雑音が重畳されたときに、雑音音声信号 $\hat{s}$ は次のようにモデル化される。

$$\hat{s} = F(s, n, \text{SNR})$$

s, n, SNR はそれぞれクリーンな音声信号、重畳する雑音、信号対雑音比を表す。 F はケプストラム空間では一般に非線形変換になる。この問題に対応するために、ケプストラム空間での確率分布を表すHMMに対して、HMM合成法[2]やneural network法[3]などの種々の非線形処理が研究されてきた。しかし、これらの手法には二つの欠点がある。①複雑な処理と大きな計算量を必要とする。②雑音 n 及び信号対雑音比 SNR が常に変動するような雑音音声に対応するのが難しい。これに対し、我々は非線形処理を区分線形変換で近似して、モデルの尤度最大化をはかる方法を提案した。さらにHMMパラメータ空間(雑音が重畳した音声のHMM空間)を雑音の性質と SNR によって木構造に区分化し、入力音声の条件に最も適合した部分空間を選ぶ方法[1]を提案した。選ばれた空間で、尤度がさらに最大化するように線形変換(MLLR[4])を行う。

2.2 雑音重畳音声GMMによる木構造クラスタリング手法

以前の手法[1]ではHMMパラメータ空間の木構造は雑音の木構造クラスタリング結果を用いていた。雑音のクラスタリングは雑音のスペクトルの距離に基づいて行うのに対し、HMMパラメータ空間は雑音重畳音声のスペクトルで表される。このためクラスタリング段階の雑音スペクトルと認識段階の雑音重畳音声のスペクトルの間に不一致が生じる。また、入力音声を確認する時、 SNR と雑音の種類二段階の探索を行うことが必要であった。即ち、まず各 SNR 条件で尤度最大なモデルを選択し、次にあらゆる SNR 条件の最適なモ

デルを比較して、その中の一番良いモデルを全体の最適なモデルとしていた。このためモデル選択の計算量が多くなる。そこで本報告では雑音重畳音声をクラスタリングする手法を提案する。この方法では雑音重畳音声の木構造クラスタリングの結果に基づいてHMMパラメータ空間の木構造を構築する。クラスタリング段階と認識段階で一貫して雑音重畳音声を用いる方法である。また現在は SNR ごとに木を構築しているが、今後は雑音重畳音声のクラスタリングによって、すべての SNR 条件のあらゆる雑音重畳音声を一つの木にすることも可能になり、入力音声を確認する時に一段階の探索で最適なモデルを選択することも可能になる。

2.3 クラスタリングの作成

雑音データは電子協雑音データベース[5]の30種類の雑音を用いた。そのうちの28種類を学習に、ほかの2種類(駅のコンコース "eki_1", 百貨店のエレベータ "hyakka_1")をテストに用いた。クリーンな音声に種々の SNR で各雑音を重畳させた後、CMSを適用した。Baum-Welchアルゴリズムを用いて64混合の各雑音重畳音声GMMを学習した。雑音重畳音声GMM間の距離を計算し、 SNR ごとに距離行列(28行28列)に基づいて雑音重畳音声のクラスタリングを行った。

3. 認識実験

3.1 音響モデル

実験で利用する音響モデルは、「話し言葉工学」プロジェクト[7]で作成された音声コーパス中の男性話者による338講演、約59時間分の音声データを用いて作成した2,000状態16混合の状態共有型triphone HMMである[8]。音声特徴量は、25次元のMFCCベクトル系列を用いた。

3.2 言語モデル

この実験は対話システムでの音声入力による飲食店舗検索とする。ユーザーは場所(最寄り駅)、料理の種類などの検索条件を発声することにより、希望の店舗の詳細情報を得る。発話内容を受理する言語モデルは、それぞれの発話内容用のテキストコーパスから作成される。クラスタ言語モデルを用い、クラスに属する単語は全て等確率で生起すると仮定する[9]。

3.3 評価用データ

システムの評価は話者10名の50発話の対話音声で行った。3種類の SNR ($\text{SNR}=5, 10, 15\text{dB}$)で、学習に用いなかった2種類の雑音 ("eki_1", "hyakka_1") 計6種類の組合せを重畳させたデータを用意した。

4. 認識実験結果

まず入力音声 SNR=15dB の場合の2種類のテストデータに対し、実験を行った。尤度最大基準に基づいて、木構造のルートから下にたどり最適なモデルを選択し、認識に用いる。クリーンHMMの結果(“Baseline”), 雑音の木構造クラスタリングを用いた場合(“従来法”), 雑音重畳音声の木構造クラスタリングを用いた場合(“提案法”)の単語正解精度(ACC%)を図1に示す。“eki_1”のテスト音声に対し、“提案法”は“従来法”と同じ結果になったが、“hyakka_1”のテスト音声に対し、“提案法”は“従来法”より高い認識率が得られた。

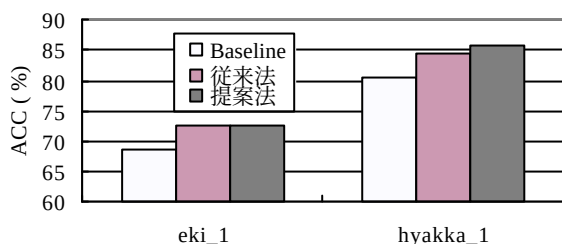


Fig.1: Comparison of noise clustering (previous) and noisv speech clustering (new) methods (SNR:15dB)

次に、50文ごとに木構造の各階層における選択結果を調べた。その結果、“eki_1”のテストデータのほとんどの文に関しては図3で示すような木構造の真中付近のモデルが選択された。図2にはテストの雑音を含めた30種類の雑音重畳音声GMMに対する数量化第4類分析[6]の結果(SNR=15dBの場合)を表す。テストとしての“eki_1”と“hyakka_1”はこの中の変数3と変数8にそれぞれ対応している。“eki_1”に対して選ばれたノードに用いられた学習雑音は駅の通路, 空調機, 仕分け工場の三種類からなっている。図2ではそれぞれ変数4, 6, 7に対応し, “eki_1”のテスト雑音に近い学習の雑音重畳音声を選択されたことが確認された。“hyakka_1”に関しても同様の結果が得られた。このことから雑音重畳音声の木構造クラスタリングの有効性が確認された。

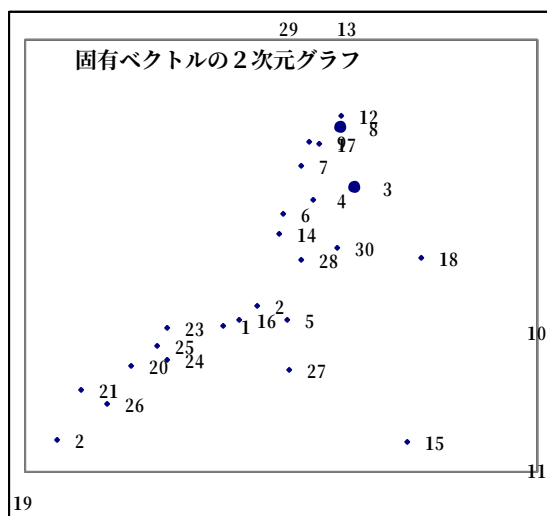


Fig.2: Analysis by Hayashi's quantification theory (type4) for noisy speech

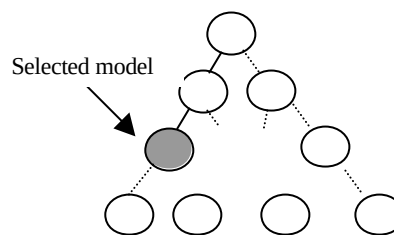


Fig.3: Model selection in the tree-structure of noise-added speech HMMs (eki_1 noise-added speech)

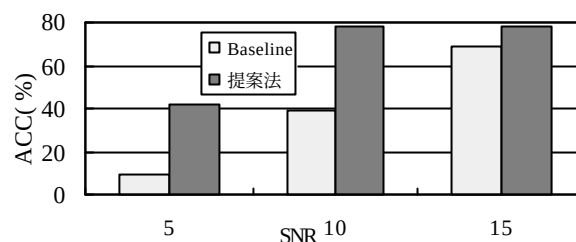


Fig.4: Recognition results (“eki_1” noise-added speech)

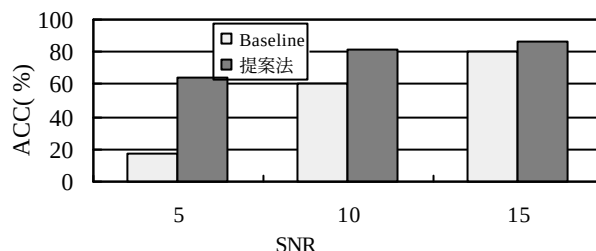


Fig.5: Recognition results (“hyakka_1” noise-added speech)

次に、SNR=5,10,15dBの全部で6種類のデータに対し、最適なノードを選択し、尤度がさらに最大化するように線形変換をする実験を行った。クリーンなHMM(“Baseline”)と提案法の単語正解精度(ACC%)を図4,5に示す。ベースラインに比べ、6種類のデータの平均で単語誤り率は36.1%低下した。

5. まとめ

尤度最大化基準に基づく区分線形雑音適応法において、雑音重畳音声GMMによるクラスタリング手法を提案した。二種類の実環境での雑音重畳音声に対して、提案手法の効果を確認した。数量化四類分析により雑音クラスタリングの効果を検討した。今後の課題には、学習雑音数の増加、木構造クラスタリングの改善、雑音区間の自動切り出しなどがある。

謝辞

討論いただいた東工大古井研究室の方々に感謝する。

参考文献

- [1] 張, 古井, 秋季音講論, pp.35-36, 2002.
- [2] F.Martin et al., 信学技報 SP92-96, 1992.
- [3] 張, 古井, 春季音講論, pp.55-56, 2001.
- [4] C.J.Leggetter et al., Computer Speech and Language, Vol.9, pp.171-185, 1995.
- [5] http://www.milab.is.tsukuba.ac.jp/corpus/noise_db.html
- [6] 菅, <<多変量解析の実践>>, 現代数学社, 1993.
- [7] 古井, 前川, 井佐原, 日本音響学会誌, vol.56, no.11, pp.752-755, 2000.
- [8] 篠崎, 古井, 秋季音講論, pp.87-88, 2002
- [9] 田熊, 森山, 岩野, 古井, 秋季音講論, pp.79-80, 2002.